Journal of the
Royal Statistical Society
SERIES A
Statistics
in Society

J. R. Statist. Soc. A (2015)

Inference and forecasting in the age–period–cohort model with unknown exposure with an application to mesothelioma mortality

María Dolores Martínez Miranda,

University of Granada, Spain, and Cass Business School, London, UK

Bent Nielsen,

Nuffield College, Oxford, UK

and Jens Perch Nielsen

Cass Business School, London, UK

[Received March 2013. Revised October 2013]

Summary. It is of considerable interest to forecast future mesothelioma mortality. No measures for exposure are available so it is not straightforward to apply a dose–response model. It is proposed to model the counts of deaths directly by using a Poisson regression with an age–period–cohort structure, but without offset. Traditionally the age–period–cohort is viewed as suffering from an identification problem. It is shown how to reparameterize the model in terms of freely varying parameters, to avoid this problem. It is shown how to conduct inference and how to construct distribution forecasts.

Keywords: Age–period–cohort; Forecasting; Identification; Mesothelioma; Mortality; Unknown exposure

1. Introduction

Standard mortality studies rest on dose–response analyses where information on both deaths and exposure is available. When no measure for exposure is available this approach is complicated. Mesothelioma mortality is one such example. Mesothelioma is a lung cancer that is almost always associated with exposure to asbestos. The usage of asbestos has been regulated for decades, yet the annual number of mesothelioma deaths continues to increase owing to the long latency of the disease. It is of considerable interest to forecast the timing of this peak as well as the overall burden of future mesothelioma mortality. Notably this is a question relating to the unconditional mortality distribution as opposed to questions relating to the distribution of mortality conditionally on survival to a certain point. There are two approaches to the problem of not having exposure data. The first approach is to construct a synthetic measure for exposure in which case a dose–response model can be used. This has been done for UK data in Peto *et al.* (1995), Hodgson *et al.* (2005) and most recently by Tan *et al.* (2010). Tan *et al.* (2010) used a Markov chain Monte Carlo method, which allows not only estimating the model parameters, but also the derivation of Bayesian credibility intervals. The second approach, which is

Address for correspondence: Bent Nielsen, Nuffield College, Oxford, OX1 1NF, UK.
E-mail: bent.nielsen@nuffield.ox.ac.uk

 0964–1998/15/178000

followed here, is to model only the responses. This is inspired by the chain ladder analysis that is used for forecasting liability reserves in general insurance; see England and Verrall (2002). The advantage of the second approach is that conceptually it is very simple as it avoids any need for constructing exposure measures for the sample array as well as any need for extrapolating the exposure into the future. We show how estimation, hypothesis testing, point forecasting and model specification testing can be done by standard frequentist statistical software. In addition, we show how distribution forecasts can be constructed by using asymptotic theory. The first approach may possibly give more accurate forecasts when expert knowledge is available. Even so, the second approach may serve as a benchmark, especially in situations where there may be some doubt about the assumptions about the unobserved exposure.

It is thought that mesothelioma is almost always caused by exposure to asbestos. It has a long latency period and it mainly affects men. Once discovered it is rapidly fatal, with most of those affected dying within 1 year; see Peto *et al.* (1995) and UK Asbestos Working Party (2009) for further details. These circumstances contribute to the accuracy of records of mesothelioma mortality and the problems in finding reliable measures on exposure as well as data on mortality from competing risks. Even though malignant mesothelioma aetiology is well understood, the worldwide incidence continues to climb (see Robinson (2012)). Whereas mesothelioma-related death seems to have peaked in the USA (see Price and Ware (2009)), projections indicate that the peak has not yet been seen in other developed areas (see Clements *et al.* (2007), Peto *et al.* (1995) and Segura *et al.* (2003) for Europe and Myojin *et al.* (2012), Segura *et al.* (2003) and Park *et al.* (2012) for Asian insights). Although all these important works use sophisticated modelling techniques developed for the problem at hand, there is of course always the risk that such detailed modelling increase the modelling error leading to worse projections than simple methods. The purpose of this paper is to provide a simple benchmark method that should be helpful when checking the robustness of other more sophisticated methods.

The methodology of the analysis presented is based on an age–period–cohort model. The age–period–cohort model has a well-known identification problem making it impossible to identify the levels and growth rates of the age, period and cohort effects. Traditionally this is solved by making *ad hoc* choices of these levels and growth rates. This does not have any effect on the fit of the model but it can have a detrimental effect on forecasts; see Holford (1985) and Kuang *et al.* (2008a, b). Instead it was suggested in Kuang *et al.* (2008a, b) to use a parsimonious, freely varying parameterization of the likelihood which permits the use of standard statistical methodology. The analysis was aimed primarily at age–cohort arrays. Building on that the methodological contribution of the present paper is as follows. First, the age–period–cohort model for age–cohort data arrays is carried over to age–period data arrays. A subtle difference in the results is demonstrated and discussed. Secondly, it is shown how to conduct inference in age–period–cohort models without exposure measures. The inference of particular interest with mesothelioma data is whether the period effect is significant. Thirdly, it is shown how to make point forecasts, including the use of intercept corrections for making the forecasts robust. Fourthly, it is discussed how to make distribution forecasts. It should be noted that the results concerning identification, estimation and point forecasting will transfer in a straightforward way to situations where exposure is measured, whereas the results concerning inference and distribution forecasts depend on the model chosen.

The main empirical contributions consist of forecasts for the future burden of mesothelioma that complement the analysis by previous researchers in various ways. First, the analysis points towards a peak in 2018 of about 2094 deaths with 95% confidence interval (1978, 2210). This is less than the peak of 2700 deaths in 2020 that was predicted by Peto *et al.* (1995) using data until 1991; it is later and more severe than the peak of 1846 in 2013 that was predicted by Hodgson

et al. (2005), using data until 2001, but more in line with the peak of 2038 in 2016 that was predicted from data until 2006 by Tan *et al.* (2010); see also Tan and Warren (2009).

A more subtle empirical finding arises from a recursive analysis of the data. Using data until 1991, 2001 and 2007 the timing of the peak is robust whereas the size of the peak is gradually lowered with the arrival of new data. This gives some indication that preventive legislation from 1969 has moderated the mesothelioma epidemic. Following on from this observation the forecast is broken down by cohort group. This shows that the cohorts of youngest ages with very few observed cases contribute substantially to the future uncertainty. Most of this uncertainty can safely be ignored, since these youngest cohort groups have largely not been exposed to asbestos.

The paper is organized as follows. In Section 2 we describe the mesothelioma mortality problem which motivates the paper. This presents the intuition of the methods that are suggested in the paper and also an advance of the main empirical results. In Section 3 it is discussed how the mortality can be modelled in this situation where no information on exposure is available. In Section 4 we define the statistical model that we consider and discuss the role of exposure. The full data analysis using our methodology is described in Section 5. A discussion of the results and some conclusions are provided in Section 6. Appendix A includes technical details of the methodology proposed.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Motivation: forecasting mesothelioma mortality

We suggest a new method for inference and forecasting which does not require known exposure. This is useful for an application such as mesothelioma mortality where the number of people who were exposed to asbestos is unknown. This can serve as a relatively simple benchmark for models with constructed exposure measures. In this section we present the data and the problem which motivates the paper and we give a brief overview of the analysis proposed.

The most recently available data from UK Health Service Executive consist of annual aggregated counts of deaths in Great Britain by age for the period 1967–2007. We focus on mesothelioma deaths of males in the age range to 25–89 years owing to sparsity in the more extreme age groups. Thus, the data are an age–period array with $I = 65$ age levels and $J = 41$ periods. The total number of deaths is 31 902, with the annual observed number of deaths peaking at 1774 in 2007.

Figs 1(a)–1(c) show summary plots of observed deaths by age, period and cohort. Fig. 1(d) shows log-cumulative-deaths by 5-year age and cohort group. The curves are nearly parallel, which could be approximately captured by an age–cohort model. This is in line with the analysis of for instance Peto *et al.* (1995). However, since the curves are not exactly parallel, it is important to validate the age–cohort by testing it against the age–period–cohort model. Such a test will also be useful for other applications.

The numbers of deaths are concentrated among those in late working life and early retirement, with the number of deaths exceeding 1000 for each of the age groups 64–77 years. At the same time the table of deaths is thin at the edges in that 52 and 128 fall in the youngest age groups 25–34 and 35–39 years respectively and four and 15 fall in the oldest cohort groups 1878–1882 and 1883–1887 respectively, and 15 fall in the youngest cohort group 1967–1982. It is therefore clear from the outset that in the context of an age–period–cohort model the youngest age effects and the youngest and oldest cohort effects will be poorly estimated.

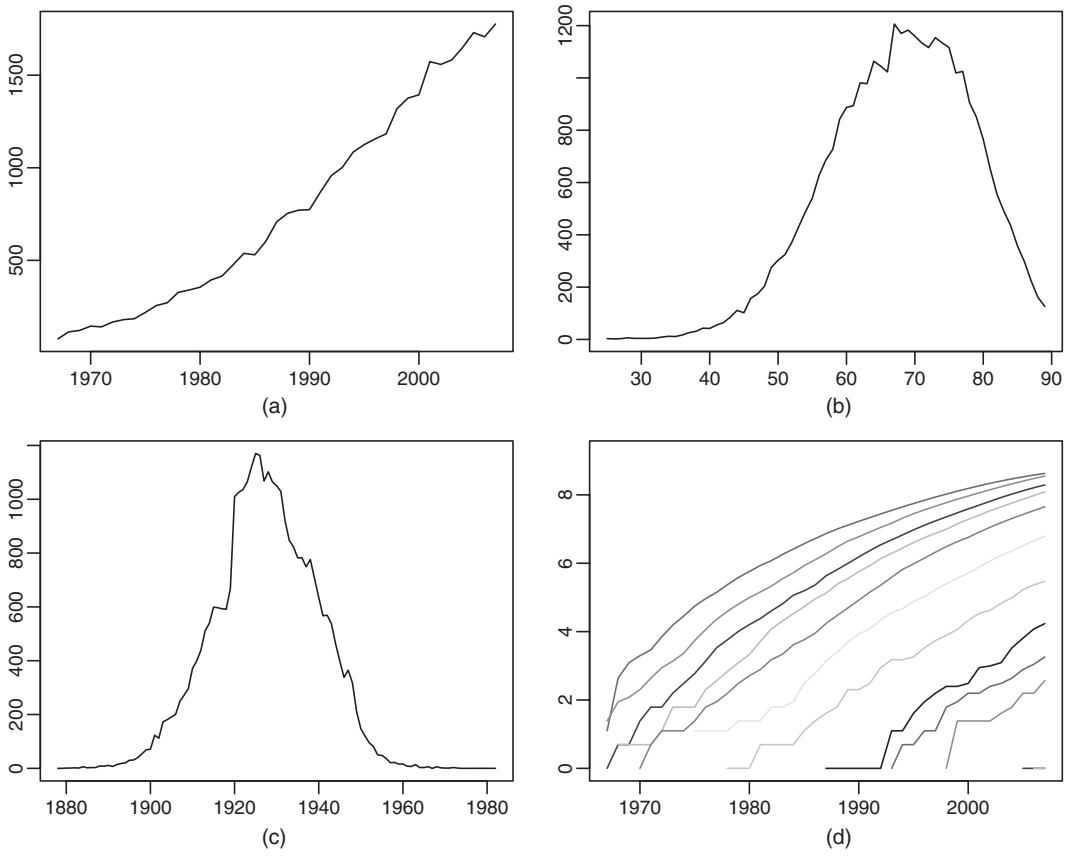

Fig. 1. Observed deaths by (a) period, (b) age and (c) cohort, and (d) log-cumulative deaths by 5-year age and cohort group: viewing the curves from top left to bottom right they represent the cohorts 1923–1927, 1928–1932, etc.

Traditionally mortality forecasts evolve around mortality ratios. In the present context no measures of exposure are available so mortality ratios are not straightforward to construct. Since the purpose of the present analysis is merely to forecast future counts of mesothelioma deaths, we shall seek to analyse the number of deaths directly by using an age–period–cohort model and then investigate various submodels through testing theory. It turns out that when exposure is unavailable the testing theory is also not available. It is therefore a contribution of this paper to derive a formal testing theory. This allows us to test the relevance of, for instance, the period effect.

The use of asbestos has been regulated in the UK since 1969 so the later cohorts in the sample have had very little exposure to asbestos. Owing to the long latency of mesothelioma this intervention is not directly visible in the data as plotted in Fig. 1. Since the main substantial question relates to asbestos-induced mortality there is less interest in analysing the younger cohorts. This has two convenient consequences. First, the data are sparse for those in the younger cohorts. We shall analyse how the data for the younger cohorts influence the analysis, but ultimately we shall settle for an analysis that excludes those sparse data. Secondly, it seems a reasonable assumption that future cohorts will largely be spared from asbestos-induced forecasts. There will therefore not be a need for extrapolating the estimated cohort parameters.

3. The role of exposure

The main statistical issue in studying asbestos-related mortality is that the exposure is unknown. Before discussing the model proposed it is useful to start with a brief review of the case where measures for exposure are available. Throughout the paper we consider data organized in an age–period array with age index $i = 1, \ldots, I$ and period index $j = 1, \ldots, J$. The cohort index $k = I - i + j$ then runs from 1 to $K = I + J - 1$.

3.1. Dose–response modelling with known exposure

First we consider the situation where both the number of deaths Y_{ij} and the exposure Z_{ij} are known across the age–period index array; see Vandenbroucke and Pearce (2012) for a general discussion. It is common to model the rates Y_{ij}/Z_{ij} by using a log-linear model This involves the assumption that the responses given exposures are Poisson distributed with expectation

$$E(Y_{ij}|Z_{ij}) = \exp(\nu_{ij})Z_{ij} = \exp\{\nu_{ij} + \log(Z_{ij})\}, \tag{3.1}$$

When ν_{ij} has a linear age–period–cohort structure the statistical model can be estimated through a Poisson regression with offset $\log(Z_{ij})$.

When forecasting the future number of deaths it is necessary to extrapolate both the model for the rates as well as the exposure. This is particularly useful if all structural changes to the mortality can be ascribed to exposure, in which case the model for the rates can be used out of sample without alterations. However, there appears to be some natural exposure to asbestos; see Tan *et al.* (2010). The asbestos legislation will therefore have changed the relative occurrence of these exposures. If mesothelioma latencies differ according to the type of exposure it is less obvious that the rates model is invariant to the changes in exposure.

3.2. Modelling mortality with synthetic exposure measures

When the exposure is not recorded the first approach is to construct synthetic exposure measures. Indeed, Peto *et al.* (1995) measured exposure in terms of the number of people in the population. This allowed the construction of rates and log-linear modelling. It was found that it was adequate to assume a simple age–cohort structure for the mortality, for $k = I - i + j$:

$$\nu_{ij} = \phi_i + \psi_k. \tag{3.2}$$

In a second analysis Hodgson *et al.* (2005) replaced the simple age–cohort structure with a more complicated model based on epidemiological insight. The log-linear model was replaced by a multinomial model for the responses Y_{ij}. The response probabilities were modelled according to a clearance model involving the half-time of clearing of asbestos fibres from the lungs and a model for exposure depending on period. The analysis was updated with more recent data in a Bayesian set-up by Tan *et al.* (2010); see also Tan and Warren (2009).

3.3. Modelling mortality without exposure

In contrast with the above approach our suggested methodology avoids the need for taking a view on exposure. As a statistical model the responses Y_{ij} are independent over the age–period array and Poisson distributed with expectation

$$E(Y_{ij}) = \exp(\mu_{ij}). \tag{3.3}$$

The predictor is assumed to have age–period–cohort structure

$$\mu_{ij} = \alpha_i + \beta_j + \gamma_k + \delta. \tag{3.4}$$

The statistical model is estimated by Poisson regression without any need for an offset. It is therefore slightly simpler than the traditional log-linear model for rates.

It is worth noting that this model concerns the marginal distribution of the responses Y_{ij} and it makes no assumptions about exposure. Legislative changes will therefore show up directly in the model for the marginal responses. This must be kept in mind when forecasting. The considerations are not all that different from the discussion for exposure in Section 3.1.

It is of considerable interest to compare the model without exposure (3.4) with the model with exposure (3.1). This can be done under simplistic assumptions about the log-expectation of the rates and about the exposure. Indeed, suppose that the log-expectation of the rates has the age–cohort structure (3.2) so $\nu_{ij} = \phi_i + \psi_k$, and that the exposure has an age–period–cohort structure $\log(Z_{ij}) = a_i + b_j + c_k + d$. In combination the log-expectation of the rates in equation (3.1) then becomes

$$\log\{E(Y_{ij}|Z_{ij})\} = \phi_i + a_i + b_j + \psi_k + c_k + d. \tag{3.5}$$

Comparing equations (3.4) and (3.5) it is seen that the rate parameters ϕ_i and ψ_k can be identified when a_i and c_k are absent. More generally, in so far as a clearance model can be formulated as a simple functional form for a_i and c_k then the parameters ϕ_i and ψ_k can be identified up to that functional form.

We shall refrain from building a model for exposure. Essentially the risk set is built up as follows. It consists of those who have survived to the time of exposure and who have then been exposed. Typically a long latency period then follows where competing risks are prevalent. Once the disease has been discovered it is rapidly fatal with less scope for mortality from competing risks. Getting to grips with the details of this development is important, especially when it comes to prevention of mesothelioma deaths. However, when the objective is merely to forecast aggregated mortality it suffices to note that this sketch of exposure indicates that the count mesothelioma mortality can plausibly be modelled as Poisson with an age–period–cohort structure. In any case, this claim can be investigated empirically. Moreover, the simplicity of the Poisson model may be advantageous in forecasting.

4. The statistical model and its analysis

An age–period–cohort model for the mortality counts is now presented and analysed. First we discuss the parameterization and show how the traditional age–period–cohort identification problem can be addressed. The model and likelihood are presented along with a proposal for inference based on a multinomial sampling scheme. Finally, we briefly present the forecast methods. Details are given in Appendix A.

4.1. Parameterization

Here, we describe how to parameterize an age–period–cohort model with a view to addressing the identification problem. A log-odds interpretation of the parameterization suggested follows. Finally, the parameterization of an age–cohort model is discussed.

4.1.1. Age–period–cohort parameterization of age–period data arrays

The age–period–cohort predictor $\mu_{ij} = \alpha_i + \beta_j + \gamma_k + \delta$ from equation (3.4) has a well-known identification problem. We can rewrite it as

$$\mu_{ij} = \alpha_i + a + id + \beta_j + b - jd + \gamma_k + c + kd + \delta - a - b - c - Id, \tag{4.1}$$

for any real numbers a, b, c and d. This shows that the time effects for age, period and cohort are identified only up to an arbitrary linear trend. By saying that the linear trend is arbitrary we mean that no method can be found to estimate that linear trend from the data. To be more precise define the time effects

$$\theta = (\alpha_1, \ldots, \alpha_I, \beta_1, \ldots, \beta_J, \gamma_1, \ldots, \gamma_K, \delta)' \in \mathbb{R}^q,$$

where $q = I + J + K + 1 = 2(I + J)$, and let μ represent the age–period array of predictors μ_{ij}. We can then find two different time effects, $\theta^\dagger \neq \theta^\ddagger$, that result in the same predictor, $\mu^\dagger = \mu^\ddagger$. This will show up as a collinearity problem in the context of a generalized linear model. In the literature there are many different *ad hoc* solutions for estimating θ. But, really, the problem is that the model is overparameterized and the solution is to find an identified parsimonious parameterization.

Corollary 2 of Kuang *et al.* (2008a) gives such a parameterization. This parameterization has two parts. One part consists of second differences of the parameters such as $\Delta^2\alpha_i = \Delta\alpha_i - \Delta\alpha_{i-1}$, where $\Delta\alpha_i = \alpha_i - \alpha_{i-1}$. The second differences are identified because the second difference of a linear trend is 0. This was pointed out by Osmond and Gardner (1982) and Clayton and Schifflers (1987). The other part consists of three points μ_{ij} chosen to pin down the shared level and linear trend. Thus, the identified parameter is

$$\xi = (\mu_{I1}, \mu_{I1} - \mu_{I-1,1}, \mu_{I2} - \mu_{I1}, \Delta^2\alpha_3, \ldots, \Delta^2\alpha_I, \Delta^2\beta_3, \ldots, \Delta^2\beta_J, \Delta^2\gamma_3, \ldots, \Delta^2\gamma_K)' \in \mathbb{R}^p, \tag{4.2}$$

where $p = q - 4 = 2(I + J - 2)$. The parameter ξ is identified in the sense that two different parameters $\xi^\dagger \neq \xi^\ddagger$ result in two different predictors $\mu^\dagger \neq \mu^\ddagger$. We shall use the notation $\xi^{(2)}$ for the last $p - 1$ components of ξ.

The next step is to find the design matrix linking the predictor μ with the parameter ξ. The design matrix depends on the data array chosen. There are three main types of data array, which were referred to as the three principle sets of dead by Lexis; see Keiding (1990). These are age–cohort arrays, age–period arrays and cohort–period arrays. The design matrices are different for these arrays in a fundamental way. Age–cohort arrays have the easiest design matrix, which is discussed in theorem 1 of Kuang *et al.* (2008a). Theorem 1 gives the link for age–period arrays, which is needed for this study. Cohort–period arrays can be dealt with in a similar way.

Theorem 1. Consider an age–period data array, $i = 1, \ldots, I$, $j = 1, \ldots, J$, along with an age–period–cohort predictor of the form (3.4), where the cohort is $k = I - i + j$. Then

$$\begin{aligned}\mu_{ij} = {} & \mu_{I1} + (i - I)(\mu_{I1} - \mu_{I-1,1}) + (j - 1)(\mu_{I2} - \mu_{I1}) \\ & + \sum_{t=i}^{I-2} \sum_{s=t}^{I-2} \Delta^2\alpha_{s+2} + \sum_{t=3}^{j} \sum_{s=3}^{t} \Delta^2\beta_s + \sum_{t=3}^{k} \sum_{s=3}^{t} \Delta^2\gamma_s.\end{aligned} \tag{4.3}$$

The parameter ξ of equation (4.2) is maximal invariant to the transformations that are listed in equation (4.1). Indeed, it is invariant and it exactly identifies μ in that $\xi^\dagger \neq \xi^\ddagger$ implies that $\mu^\dagger \neq \mu^\ddagger$.

Formula (4.3) has an interesting interpretation. We shall highlight a few aspects of this. First, formula (4.3) writes the predictor in terms of one overall level, two linear trends and three time effects. An *ad hoc* identification would allocate these two linear trends to the three time effects in some arbitrary way. There are such examples in the literature. An early, popular method in cancer research is that of Osmond and Gardner (1982) and Gardner and Osmond (1984).

Ad hoc identification will of course not add anything to the statistical analysis, apart from an arbitrariness which, at best, will not disturb the statistical analysis. The difficulty is that the co-ordinates of the *ad hoc* identified time effects do not vary freely in contrast with both the unidentified time effects that were initially used to construct the model and the canonical parameter ξ. Interpretation and time series analysis of the *ad hoc* identified time effects are therefore easily driven by arbitrary identification choice. This can sometimes happen in ways that are not entirely obvious. Kuang *et al.* (2008b) have discussed such issues in relation to forecasting.

Secondly, an interesting feature of formula (4.3) is that period and cohort differences are cumulated forwards, whereas the age effect is cumulated backwards. This is because the principal axes are age and cohort, so period arises as a difference. It is therefore not possible to choose a reference point from which all three timescales increase. This is different for age–cohort arrays where the period increases with age and cohort as explored in theorem 1 of Kuang *et al.* (2008a).

Thirdly, as we shall now show, the double differences $\Delta^2\alpha_i$, $\Delta^2\beta_j$ and $\Delta^2\gamma_k$ can be interpreted as log-odds ratios whereas the slopes $\mu_{I1}-\mu_{I-1,1}$ and $\mu_{I2}-\mu_{I1}$ can be interpreted as log-odds.

4.1.2. *A log-odds-ratio interpretation of the parameterization*

The canonical parameter can be given a log-odds-ratio interpretation. To see this introduce the aggregate mean and the frequencies for each cell as

$$\begin{aligned} \tau &= E(Y_{..}), \\ \pi_{ij} &= E(Y_{ij})/E(Y_{..}), \end{aligned} \tag{4.4}$$

where $Y_{..}=\Sigma_{ij}Y_{ij}$ and where $\log\{E(Y_{ij})\}=\mu_{ij}$. This predictor satisfies the identity (4.3) in theorem 1 and it therefore varies on a p-dimensional manifold in an $n=IJ$-dimensional real space. Similarly, the frequencies π_{ij} vary on a manifold of dimension $p-1$ which can be parameterized by $\xi^{(2)}$.

The frequencies are log-linear functions of the parameters $\xi^{(2)}$ which can be expressed in terms of various log-odds ratios and log-odds. Indeed, formula (4.3) in theorem 1 implies the log-odds ratios

$$\log\left(\frac{\pi_{ij}}{\pi_{i-1,j}}\Big/\frac{\pi_{i-1,j-1}}{\pi_{i-2,j-1}}\right)=\mu_{ij}-\mu_{i-1,j}-\mu_{i-1,j-1}+\mu_{i-2,j-1}=\Delta^2\alpha_i, \tag{4.5}$$

$$\log\left(\frac{\pi_{ij}}{\pi_{i,j-1}}\Big/\frac{\pi_{i-1,j-1}}{\pi_{i-1,j-2}}\right)=\mu_{ij}-\mu_{i,j-1}-\mu_{i-1,j-1}+\mu_{i-1,j-2}=\Delta^2\beta_j, \tag{4.6}$$

$$\log\left(\frac{\pi_{ij}}{\pi_{i,j-1}}\Big/\frac{\pi_{i+1,j}}{\pi_{i+1,j-1}}\right)=\mu_{ij}-\mu_{i,j-1}-\mu_{i+1,j}+\mu_{i+1,j-1}=\Delta^2\gamma_k. \tag{4.7}$$

This log-odds-ratios interpretation is similar to the relative risk interpretation that was offered by Clayton and Schifflers (1987). The formulae are illustrated in Fig. 2. Fig. 2(a) illustrates the interpretations of the formula for $\Delta^2\alpha_i$ as follows. Consider the 1970 and 1971 cohorts. In 2010 these have age 40 and 39 years, whereas in 2011 these have age 41 and 40 years. Thus, $\Delta^2\alpha_{41}$ represents the increase in mortality from age 40 to age 41 years in 2011 relative to the increase from age 39 to age 40 years in 2010. In a similar way Figs 2(b) and 2(c) illustrate the formulae for $\Delta^2\beta_{2012}$ and $\Delta^2\gamma_{1972}$.

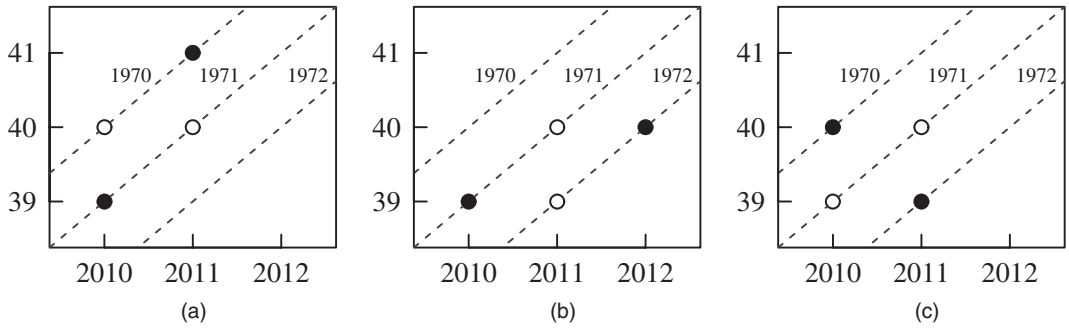

Fig. 2. Illustration of log-odds-ratio interpretation of (a) $\Delta^2\alpha_{41}$, (b) $\Delta^2\beta_{2012}$ and (c) $\Delta^2\gamma_{1972}$

The slope parameters can be interpreted as log-odds of the frequencies for the reference points $(I,1)$, $(I-1,1)$ and $(I,2)$ in that

$$\begin{aligned}\log(\pi_{I1}/\pi_{I-1,1}) &= \mu_{I1} - \mu_{I-1,1},\\ \log(\pi_{I2}/\pi_{I1}) &= \mu_{I2} - \mu_{I1}.\end{aligned} \tag{4.8}$$

These calculations of log-odds ratios and log-odds imply that there is a log-linear one-to-one mapping between the frequencies π_{ij} and the freely varying parameters $\xi^{(2)}$. Moreover, the aggregate mean parameter τ satisfies

$$\tau = \exp(\mu_{I1}) \sum_{ij} \exp(\mu_{ij} - \mu_{I1}).$$

It is therefore a product of a term that depends on the level parameter μ_{I1} and a term that depends on the remaining part of the canonical parameter, $\xi^{(2)}$. It should be noted that the log-odds ratios of frequencies are discussed here for their interpretational value, whereas likelihoods are preferably parameterized in terms of the parameter ξ. More details are given in Section 4.2.

4.1.3. Age–cohort parameterization of age–period data arrays

In the empirical analysis we shall be interested in testing the absence of a period effect. The change of model changes the identification discussion so it is worth giving the results also for this submodel. The predictor (3.4) reduces to

$$\mu_{ij} = \alpha_i + \gamma_k + \delta. \tag{4.9}$$

The two linear trends in the age–period–cohort representation (4.3) can now be attributed uniquely to the age and cohort effects, since the period effect is now absent. The identitification problem (4.1) therefore reduces to

$$\mu_{ij} = \alpha_i + a + \beta_j + b + \delta - a - b, \tag{4.10}$$

for any real numbers a and b. Likewise theorem 1 can be modified as follows.

Theorem 2. Consider an age–period data array, $i = 1,\ldots,I$, $j = 1,\ldots,J$, along with an age–cohort predictor of the form (4.9), where the cohort is $k = I - i + j$. Then

$$\mu_{ij} = \mu_{I1} - \sum_{t=i}^{I-1} \Delta\alpha_{t+1} + \sum_{t=2}^{k} \Delta\gamma_t. \tag{4.11}$$

The parameter

$$\xi_{\mathrm{AC}} = (\mu_{I1}, \Delta\alpha_2, \ldots, \Delta\alpha_I, \Delta\gamma_2, \ldots, \Delta\gamma_K)' \tag{4.12}$$

is maximal invariant with respect to the transformations listed in equation (4.10). Indeed, it is invariant to these transformations and it exactly identifies μ in that $\xi^{\dagger}_{\mathrm{AC}} \neq \xi^{\ddagger}_{\mathrm{AC}}$ implies that $\mu^{\dagger} \neq \mu^{\ddagger}$.

The interpretation of the age–cohort result is similar to that of the age–period–cohort result, albeit simpler. In analysis of variance the differences $\Delta\alpha_i$ and $\Delta\gamma_k$ would be referred to as contrasts. The contribution of equation (4.11) is the identity between the predictor and a combination of the contrasts and a common level parameter. Since the data are arranged in an age–period array it is not possible to find a single point from which both the age and the cohort array increases. Therefore age is cumulated backwards as before.

4.2. Statistical analysis

Having clarified the parameterization the statistical model can now be presented. This is followed by a discussion of the sampling scheme and hypothesis testing.

4.2.1. Statistical model

The available data are the responses Y_{ij} over an age–period array. No measures of exposure are available. Initially, we shall assume that the responses Y_{ij} are independent and Poisson distributed over the age–period array where the log-mean $\log\{E(Y_{ij})\} = \mu_{ij}$ satisfies the age–period–cohort model. Theorem 1 shows how to parameterize μ_{ij} in terms of the freely varying parameter ξ. The Poisson likelihood is then

$$\log\{L(\xi; Y)\} = \sum_{i,j} Y_{ij} \log\{E(Y_{ij})\} - \sum_{i,j} E(Y_{ij}). \tag{4.13}$$

Formula (4.3) implies that $\log\{E(Y_{ij})\} = \mu_{ij} = X'_{ij}\xi$, where

$$X_{ij} = \{1, i-I, j-1, h(1,i), \ldots, h(I-2,i), h(j,3), \ldots, h(j,J), h(k,3), \ldots, h(k,K)\}', \tag{4.14}$$

and $h(t,s) = \max(t-s+1, 0)$. Since ξ is freely varying the Poisson likelihood is a regular exponential family with ξ as canonical parameter; see Barndorff-Nielsen (1978), page 116. The Poisson likelihood is therefore maximized by a Poisson regression of Y_{ij} on X_{ij} with a log-link and no offset.

4.2.2. Multinomial sampling

To introduce a multinomial sampling scheme we rewrite the canonical parameter ξ in terms of a mixed parameterization of a mean value parameter and a part of the canonical parameter; see Barndorff-Nielsen (1978), page 121, and McCullagh and Nelder (1999), page 210. This is done in terms of the aggregate mean τ and the frequencies π_{ij} of equation (4.4). The Poisson likelihood (4.13) is rewritten by adding and subtracting $Y_{..} \log(\tau)$ to obtain

$$\log\{L(\xi; Y)\} = Y_{..} \log(\tau) - \tau + \sum_{i,j} Y_{ij} \log(\pi_{ij}). \tag{4.15}$$

The first term is the Poisson likelihood for τ based on $Y_{..}$, whereas the second term is the multinomial likelihood for the frequencies π_{ij} based on the data array Y conditionally on the sum $Y_{..}$. Exponential family theory shows that the parameters of the two likelihoods vary freely, i.e. τ on the one hand and π_{ij} on the other hand vary freely. Maximum likelihood estimation of τ and π_{ij} can be done separately from the two likelihoods.

Inference will be conducted by using a multinomial sampling scheme, i.e. asymptotic theory will be based on a large value of the aggregate mean parameter τ. This sampling scheme permits

inference on the frequencies π_{ij} or, equivalently, on $\xi^{(2)}$, which are the $p-1$ last elements of the canonical parameter ξ. The hypothesis that the age–period–cohort structure does not depend on period is indeed of this form.

In the application we found it easier to estimate the parameters from the Poisson likelihood (4.13) rather than equation (4.15). For the inference we then need to use the δ-method to move from the asymptotic theory for the frequencies π_{ij} to an asymptotic theory for $\xi^{(2)}$. The details are left to Appendix A.2.

4.2.3. Testing for absence of period effect

With the mesothelioma data it is of particular interest to test the absence of the period effect. The age–period–cohort predictor (3.4) then reduces to the age–cohort predictor (4.9). The canonical parameter is given by equation (4.12) whereas the design reduces to

$$X_{ij}^{\mathrm{AC}} = \{1, -h^{\mathrm{AC}}(1,i), \ldots, -h^{\mathrm{AC}}(I-1,i), h^{\mathrm{AC}}(k,2), \ldots, h^{\mathrm{AC}}(k,K)\}', \tag{4.16}$$

where $h^{\mathrm{AC}}(t,s) = \mathbf{1}_{(t \geqslant s)}$.

Explicit expressions for the maximum likelihood estimators can be established for age–cohort arrays and also for triangular arrays. In the latter case the model is known as a chain ladder model; see Kuang *et al.* (2009). For age–period arrays it does not seem easy to find analytic expressions for the estimators.

The age–cohort model can be tested against the general age–period–cohort model by using a multinomial sampling scheme. The deviance is then asymptotically χ^2 with $J-2$ degrees of freedom.

4.3. Forecasting

In the empirical analysis the data are organized in an age–period array, $\mathcal{I}$ say, of dimension $I \times J$. Generally, it will be of interest to forecast h periods ahead. In the mesothelioma context this simplifies since the cohorts of the youngest men have not had much asbestos exposure. It is then only of interest to extrapolate those cohorts which are included in the sample. The forecast period can therefore be captured by the triangular array

$$\mathcal{J} = \{(i,j) : i = 1, \ldots, I; j = J+1, \ldots, J+h; k = 1, \ldots, K\}. \tag{4.17}$$

This is a subset of the rectangular set

$$\mathcal{K} = \{(i,j) : i = 1, \ldots, I; j = J+1, \ldots, J+h;\}. \tag{4.18}$$

Forecasting on the set $\mathcal{K}$ is done through a bivariate generalization of the methods that are presented here, which is a complexity that we shall not need for the present mesothelioma data. In what follows we discuss five different aspects of forecasting: the choice of the out-of-sample model, point forecasting, distribution forecasting, robust forecasts and finally the alternative approach of incorporating the forecast model directly in a statistical model.

4.3.1. Choice of out-of-sample model

The statistical model has been fitted only to the data indexed by $\mathcal{I}$. With the mesothelioma data there are no signs of major structural changes out of sample so the general idea is to seek to extrapolate the fitted model.

4.3.2. Point forecasting

Within the age–period–cohort model the period parameter would have to be extrapolated to forecast over the index set $\mathcal{J}$. This can be done by interpreting the estimates for the period param-

eter as data, fitting an auto-regression and using auto-regressive time series forecast techniques. The identification of the age–period–cohort model matters. In general, *ad hoc* identification will introduce arbitrary effects in the forecasts. The theory that was developed in Kuang *et al.* (2008b) gives a necessary and sufficient condition for avoiding such arbitrariness.

First, consider point forecasting from an age–cohort model over the set $\mathcal{J}$; see equation (4.17). In this case there is no need to extrapolate the estimate $\hat{\xi}$. Assume that $Y_{i,J+h}$ is Poisson$\{\exp(\mu_{i,J+h})\}$ distributed with a log-predictor that can be estimated by

$$\tilde{\mu}_{i,J+h} = \tilde{X}^{\mathrm{AC}\prime}_{i,J+h}\hat{\xi}^{\mathrm{AC}}, \tag{4.19}$$

where $\hat{\xi}^{\mathrm{AC}}$ estimates ξ^{AC} defined in equation (4.12) and the design (4.16) is extended so that

$$\tilde{X}^{\mathrm{AC}}_{i,J+h} = \{1, -h^{\mathrm{AC}}(1,i), \ldots, -h^{\mathrm{AC}}(I-1,i), h^{\mathrm{AC}}(k,2), \ldots, h^{\mathrm{AC}}(k,K)\}'. \tag{4.20}$$

The multinomial parameter is extrapolated by

$$\tilde{\pi}_{i,J+h} = \hat{\tau}^{-1}\exp(\tilde{\mu}_{i,J+h}), \tag{4.21}$$

which is positive, but not necessarily bounded by 1. The point forecast of the number of deaths is therefore

$$\tilde{Y}^{\mathrm{point}}_{i,J+h} = \exp(\tilde{\mu}_{i,J+h}) = \hat{\tau}\,\tilde{\pi}_{i,J+h}. \tag{4.22}$$

Next, consider point forecasting from an age–period–cohort model over the set $\mathcal{J}$; see equation (4.17). In this case the estimate $\hat{\xi}$ must be extrapolated. Form a time series $x_3, \ldots, x_J$ by

$$x_j = \sum_{t=3}^{j}\sum_{s=3}^{t}\Delta^2\beta_s$$

using the canonical parameter $\breve{\xi}$ of equation (A.1) in Appendix A. Fit the linear regression

$$x_j = \nu_c + \nu_l j + \varepsilon_j \qquad \text{for } j = 3, \ldots, J \tag{4.23}$$

by least squares. This gives the extrapolation

$$\tilde{x}_{J+h} = \hat{\nu}_c + \hat{\nu}_l(J+h), \tag{4.24}$$

which is constructed in terms of $\bar{j} = (J-2)^{-1}\sum_{j=3}^{J} j$ and $\bar{x} = (J-2)^{-1}\sum_{j=3}^{J} x_j$ and the estimators

$$\hat{\nu}_l = \sum_{j=3}^{J} x_j(j-\bar{j}) \Big/ \sum_{j=3}^{J}(j-\bar{j})^2,$$

$$\hat{\nu}_c = \bar{x} - \hat{\nu}_l\bar{j}.$$

To obtain the overall point forecast of the predictor insert this in equation (4.3):

$$\tilde{\mu}_{i,J+h} = \hat{\mu}_{I1} - (I-i)(\hat{\mu}_{I1} - \hat{\mu}_{I-1,1}) + (j-1)(\hat{\mu}_{I2} - \hat{\mu}_{I1}) + \sum_{t=i}^{I-2}\sum_{s=t}^{I-2}\Delta^2\hat{\alpha}_{s+2} + \tilde{x}_{J+h} + \sum_{t=3}^{k}\sum_{s=3}^{t}\Delta^2\hat{\gamma}_s. \tag{4.25}$$

The extrapolation method (4.23) is linear trend preserving. This is not a requirement when working directly with canonical parameters. However, it becomes a requirement when working with *ad hoc* identified period effects due to the analysis of Kuang *et al.* (2008b). The *ad hoc* identified period effect is of the form $x_j^{\mathrm{adhoc}} = x_j + b + dj$ for $j = 3, \ldots, J$ and $x_j^{\mathrm{adhoc}} = b + dj$ for $j = 1, 2$, for some arbitrarily chosen real values b and d. Applying this method to x_j^{adhoc} for $j = 3, \ldots, J$ rather than to x_j will give exactly the same forecast for the predictor $\mu_{i,J+h}$.

A subtle point is that the linear trend forecast could also be applied to x_j^{adhoc} for $j=1,\ldots,J$. This forecast is, however, different. It is linear trend preserving so the difference is not arising from arbitrariness of the *ad hoc* identification, but rather arising from the different time series properties of the two extrapolation methods. Graphing the time series x_j^{adhoc} for $j=1,\ldots,J$ will in general reveal that the first two observations stand out from the rest of the series. The only exception is when the *ad hoc* identification is chosen to avoid this, but this is rarely so. From a time series viewpoint a fit to the entire series $j=1,\ldots,J$ would therefore not seem appropriate.

4.3.3. Distribution forecasting

In a Bayesian model like that of Tan *et al.* (2010) distribution forecasting is done by simulation. In the model presented, asymptotic forecast error bands can be computed analytically by using the δ-method. The result is presented for the age–cohort model. To construct a distribution forecast the distribution of the difference between the eventual outcome Y_{ij} and the point forecast $\tilde{Y}_{i,J+h}^{\text{point}}$ must be assessed by using the multinomial sampling scheme. Therefore write

$$\tau^{-1/2}(Y_{i,J+h}-\tilde{Y}_{i,J+h}^{\text{point}})=\tau^{-1/2}(Y_{i,J+h}-\tau\pi_{i,J+h})-\tau^{1/2}(\tilde{\pi}_{i,J+h}-\pi_{i,J+h}). \tag{4.26}$$

There are three ingredients to the asymptotic analysis.

The first term in equation (4.26) is the Poisson-distributed innovation error. For large τ then

$$\tau^{-1/2}(Y_{i,J+h}-\tau\pi_{i,J+h})\xrightarrow{\text{D}}N(0,\pi_{i,J+h}).$$

Replace the variance $\pi_{i,J+h}$ by $\tilde{\pi}_{i,J+h}$ calculated in equation (4.21), noting that the estimation uncertainty is of order $\tau^{-1/2}$ and can be ignored for practical purposes.

The second term in equation (4.26) is the estimation error. The estimation uncertainty for $\tilde{\pi}_{ij}$ when $i,j\in\mathcal{I}$ is given in equation (A.4) in Appendix A. Similarly, for $i, J+h\in\mathcal{J}$ it holds that

$$\tau^{1/2}(\tilde{\pi}_{i,J+h}-\pi_{i,J+h})\xrightarrow{\text{D}}N(0,\pi_{i,J+h}^2 s_{i,J+h,\text{est}}^2),$$

where $s_{i,J+h,\text{est}}^2$ depends on the information for $\xi^{(2)}$. An explicit expression is given in equation (A.9).

Thirdly, the innovation error relating to the index set $\mathcal{J}$ and the estimation error relating to the index set $\mathcal{I}$ are independent because of the independence assumption.

The overall distribution forecast is then

$$\tilde{Y}_{i,J+h}^{\text{distribution}}=\hat{\tau}\tilde{\pi}_{i,J+h}+N\{0,\hat{\tau}\tilde{\pi}_{i,J+h}(1+s_{i,J+h,\text{est}}^2)\}. \tag{4.27}$$

As mentioned above, an explicit expression for $s_{i,J+h,\text{est}}^2$ is given in equation (A.9). Likewise, Appendix A.3 also discusses the construction of aggregated forecasts.

Distribution forecasting was also considered by Elkum (2005) in the context of an *ad hoc* identified age–period model. That analysis does not seem to distinguish errors from estimation and forecast innovations.

4.3.4. Robust point forecasting

When forecasting one is often faced with the problem that the model for the data does not quite extend to the forecast period. Small or large jumps in the data at the forecast origin can result in forecast failure. The lesson from the time series literature is that the main cause of forecast failure is jumps in the mean. We discuss two ways of making the extrapolation robust: intercept corrections and working with differenced data; see also Hendry and Nielsen (2007), chapter 21, for a discussion in the context of time series data.

For the mesothelioma data an intercept correction appears sensible since the last observations appear to jump relatively to the previous observation without an indication of a permanent shift in the trend. The idea is simply to add the last in-sample residual to the standard forecasts outlined above. Since the mean is parameterized on a log-scale this is done as follows. Recall the aggregated point forecast $\tilde{Y}^{\text{point}}_{\cdot,J+h}$ from equation (A.10) in Appendix A and note that the in-sample predictor for $Y_{\cdot,J}$ is $\hat{Y}_{\cdot,J} = \Sigma_{i=1}^{I} \hat{\tau}\hat{\pi}_{iJ}$. The intercept correction is then given by

$$\tilde{Y}^{\text{point,IC}}_{\cdot,J+h} = \tilde{Y}^{\text{point}}_{\cdot,J+h} Y_{\cdot,J} / \hat{Y}_{\cdot,J} = \hat{\tau} \sum_{i=h+1}^{I} \tilde{\pi}_{i,\,J+h} \sum_{i=1}^{I} Y_{iJ} \Big/ \sum_{i=1}^{I} \hat{\tau}\hat{\pi}_{iJ}. \tag{4.28}$$

For the age–period–cohort model another possibility for intercept correction arises in relation to the extrapolation of the period effect. Working with the linear extrapolation method of equation (4.24) the idea is to add the residual $\hat{\varepsilon}_J = x_J - \hat{\nu}_c - \hat{\nu}_l$ to the forecast $\tilde{x}_{J+h}$, giving the intercept-corrected extrapolation

$$\tilde{x}^{\text{IC}}_{J+h} = \tilde{x}_{J+h} + \hat{\varepsilon}_J = x_J + \hat{\nu}_l h. \tag{4.29}$$

This is then inserted in equation (4.25) instead of $\tilde{x}_{J+h}$ from equation (4.24).

When there are more abrubt changes in the time series robust forecasts near the forecast origin it can be important to make them robust against this by working with the differenced data. Such an abrubt change is seen in the data albeit in the middle of the sample. Fig. 1(a) shows the number of deaths by period. It is seen that there is a tendency to exponential growth until about 1987, after which the growth slows down. This could be a result of the asbestos legislation that was introduced from 1969 and onwards. If only data until about 1991 were available this issue would be critical for forecasting as illustrated in Section 5.5. With that restricted set of data men in the youngest cohorts would have been exposed to asbestos, but because of the long latency of mesothelioma the cohort effects would be poorly estimated. Kuang *et al.* (2011) discussed robustification of an age–period–cohort method applied to an age–cohort array of data, whereas a general time series discussion is given in Clements and Hendry (1999) or Hendry and Nielsen (2007), chapter 21. Those methods could have proved helpful if only data until 1991 had been available.

4.3.5. Direct combination of forecast model and statistical model

An alternative to the approach discussed above would be to incorporate the auto-regressive forecasting model in the statistical model either through a random-effects approach as in Lee and Lin (1996) or through a Bayesian approach as in Berzuini and Clayton (1994). We are less comfortable with those approaches for three reasons. First, when following such approaches the identification problem must be analysed carefully. In an analysis which we shall publish elsewhere we find that priors should be introduced only on the canonical parameters. Introducing priors on the full time effect parameter as done in Lee and Lin (1996) and Berzuini and Clayton (1994) effectively amounts to an *ad hoc* identification that cannot be updated by the likelihood. The posteriors and the forecasts will therefore have arbitrary elements that are not informed by the data in a similar way to the frequentist case that was discussed in Kuang *et al.* (2008b). Secondly, even though the forecast model is built into the statistical model, forecasting still requires extrapolation. Even if an in-built auto-regressive model for the period effect appears reasonable in sample, it will not in general be appropriate out of sample. Thirdly, real data are rarely as clean as one would like them to be. A simple statistical model is typically useful as it allows simple ways to do specification analysis and sensitivity analysis.

5. Analysis of the mesothelioma data

Here we provide the full data analysis for the mesothelioma data that were described in Section 2 and explain and derive in detail our conclusions advanced in Section 5.7. The steps in our data analysis are outlined as follows. The specification of the age–period–cohort model is assessed in Section 5.1. The parameter estimates for the age–period–cohort model are reported in Section 5.2. A reduction of the model to an age–cohort specification is tested in Section 5.3. Forecasts of the annual number of deaths are reported in Section 5.4. In Section 5.5 we perform a recursive forecast analysis. Thereby the forecasts suggested can be compared with previous analyses in the literature and we gain an insight into the variations that will arise as more data become available in the future. Finally, the sensitivity of the forecasts with respect to variations of the model and the data is analysed in Section 5.6.

5.1. Specification analysis

We start by fitting a general age–period–cohort model. Our first aim is to check that the model is correctly specified. Table 1 reports the deviance of the age–period–cohort model against a fully saturated model to be 2384.9 with a p-value of 0.852. Thus, we cannot reject that the model is well specified.

As a complement to this analysis, Fig. 3 gives a map of the standardized (Poisson) residuals defined by $r_{ij} = (Y_{ij} - \hat{Y}_{ij})/\sqrt{\hat{Y}_{ij}}$. The standardized residuals are asymptotically normal for large values of the expectation. Although this approximation is not ideal for the oldest cohorts and the youngest age groups, it still serves to illustrate whether there is any obvious pattern in the data that is not caught by the age–period–cohort model. The residuals are coded on a greyscale according to their absolute value. The small and large residuals appear to be scattered almost randomly apart from a slight tendency for the largest residuals to be concentrated on ages up to 45 years, which is where the data are sparse. A sensitivity analysis is therefore conducted in Section 5.6.

5.2. Parameter estimates

Owing to the identification problem of the age–period–cohort model, levels and the growth rates of the age, period and cohort effects cannot be identified from the model. We refrain from *ad hoc* identifications of these level and slope effects, since such identification will potentially distort inferences and forecasts. Instead we focus on a presentation of estimates of the canonical parameters.

The canonical parameter can be presented either in terms of double differences as in equation (4.2) or in terms of double sums of double differences as in equation (A.1) in Appendix A. Both versions are presented in Fig. 4. Figs 4(a), 4(c) and 4(e) show the estimated double differences of age, period and cohort effects respectively. They can be interpreted as log-odds ratios as

Table 1. Deviance analysis of the age–period–cohort model

Model	*Deviance*	*Degrees of freedom*	*p*
Age–period–cohort	2384.9	2457	0.852
Age–cohort	2441.7	2496	0.778
Age–cohort *versus* age–period–cohort	56.8	39	0.033

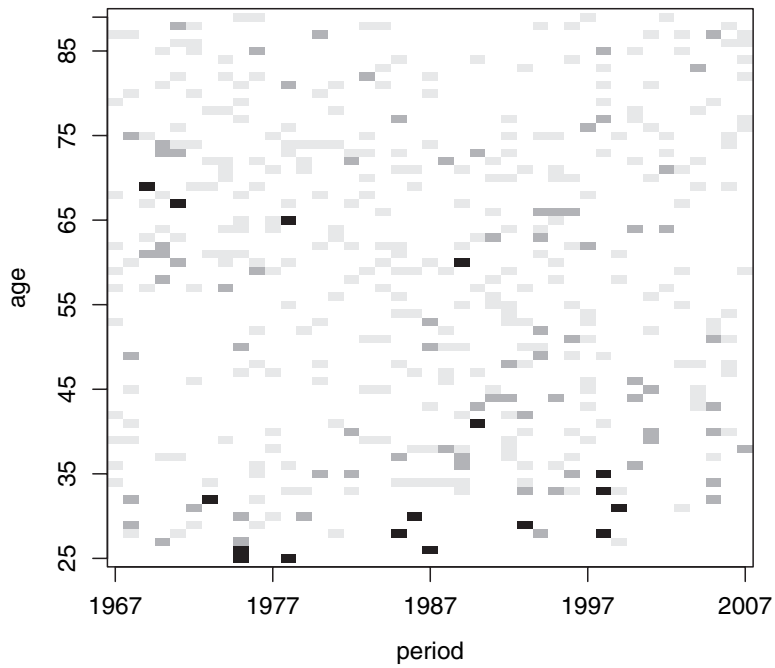

Fig. 3. Map of standardized residuals r_{ij} (Poisson, age–period–cohort model, whole data): □, $|r_{ij}| < 1$; ▒, $1 \leqslant |r_{ij}| < 2$; ▒, $2 \leqslant |r_{ij}| < 3$; ■, $3 \leqslant |r_{ij}|$

described in equations (4.5)–(4.7). We discuss their interpretation below. Figs 4(b), 4(d) and 4(f) show the estimated double sums of double differences. We shall use these to illustrate the identification issue.

The double differences that are shown in Figs 4(a), 4(c) and 4(e) are very volatile for this particular data set. It will therefore be difficult to gain intuition from these plots. The volatility is quite possibly a small sample effect. Indeed, the sparsity of the observations for the youngest age groups and the youngest and oldest cohort groups clearly shows up: the graph of the age effects, $\Delta^2\alpha_i$, in Fig. 4(a) is very volatile for young ages. The cohort effects, $\Delta^2\gamma_k$, in Fig. 4(e) suffer from a similar problem for the youngest and the oldest cohorts. In Section 5.6 it is evaluated to what extent this sparsity influences the results.

It would of course be interesting if some pattern could be found in these series of log-odds ratios. For instance, the age-related log-odds ratios would have a particularly simple interpretation if all $\Delta^2\alpha_i$s could be restricted to be equal. This in turn would imply that the overall age effect would be quadratic. Later, in Section 5.3, we shall test such patterns on these coefficients. A similar inspection of the period effects in Fig. 4(c) suggests that the double differences $\Delta^2\beta_j$ for the period effects are close to white noise so the period effect may actually be negligible. This was the working hypothesis in the dose–response analysis by Peto *et al.* (1995). Again this will be discussed further in Section 5.3.

Double sums of the estimated second differences are shown in Figs 4(b), 4(d) and 4(f). First of all, note that the age effect that is shown in Fig. 4(b) is not defined for the two oldest age groups, whereas in Figs 4(d) and 4(f) the period and the cohort are not defined for the smallest two values of the index. In Figs 4(d) and 4(f) we also illustrate the effect of *ad hoc* identification. The slope of the linear trends in the double sums of the age and the cohort effects

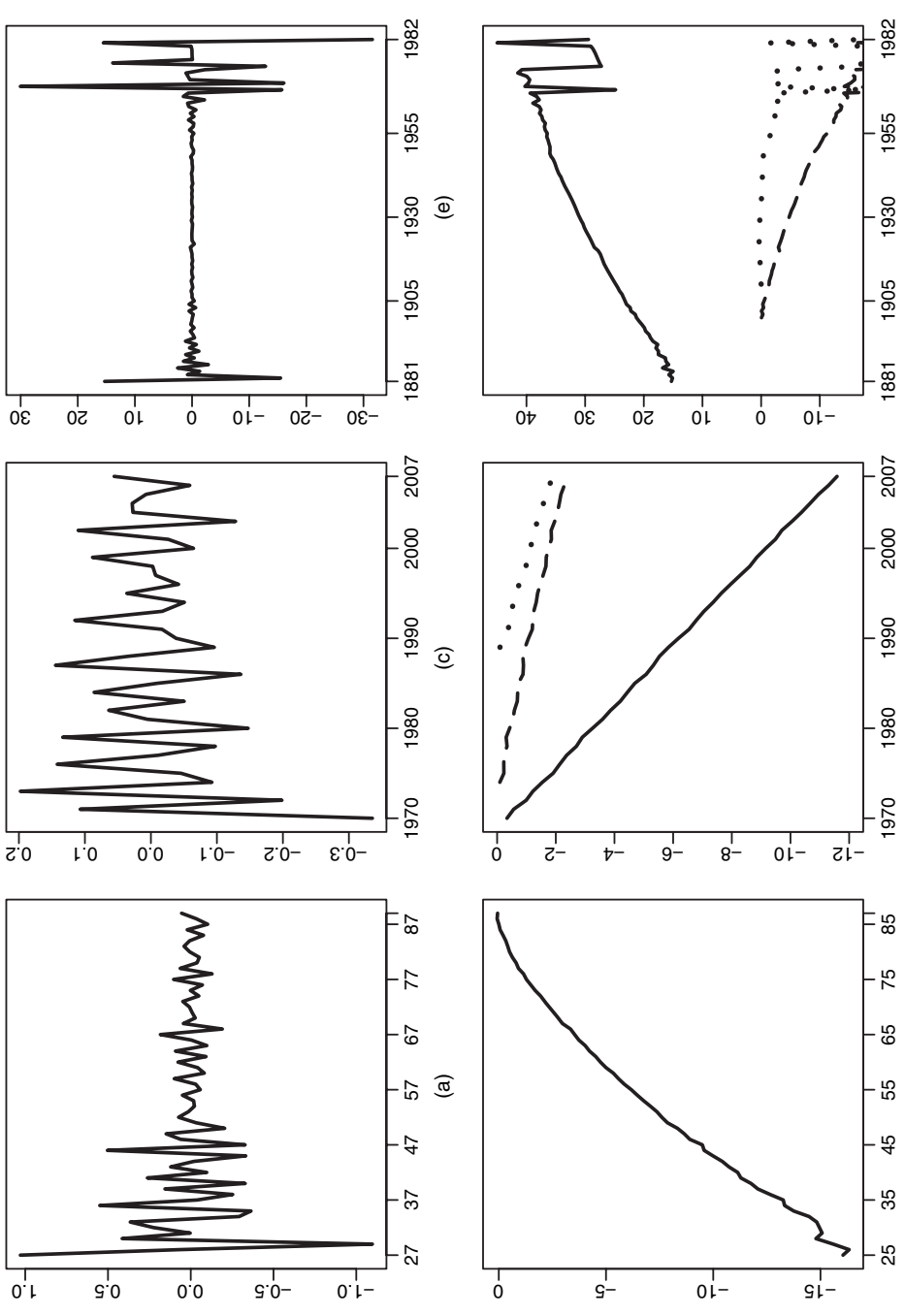

Fig. 4. Estimated effects in the age–period–cohort model: (a) age, $\Delta^2\alpha$; (b) age, $\Sigma\Sigma\Delta^2\alpha$; (c) period, $\Delta^2\beta$; (d) period, $\Sigma\Sigma\Delta^2\beta$ (——, 1969–2007; — —, 1974–2007; · · · · · ·, 1989–2007); (e) cohort, $\Sigma\Sigma\Delta^2\gamma$ (——, 1880–1982; — —, 1900–1982; · · · · · ·, 1910–1982)

are essentially driven by the first few double differences. Ignoring the first few double differences and starting the double cumulation later therefore gives quite a different appearance, because different double differences generate the linear trends. This can be interpreted as showing the effect of *ad hoc* identification. For instance, in Fig. 4(d), the bottom curve shows the effect of *ad hoc* identification by imposing that the two first age effects are 0, $\beta_1 = \beta_2 = 0$. Similarly, the next curves show *ad hoc* identification by $\beta_6 = \beta_7 = 0$ and $\beta_{21} = \beta_{22} = 0$. Here, all curves are downward sloping. Repeating the exercise for the cohorts reveals that the sign of the slope can easily change by *ad hoc* identification, which illustrates the danger of the traditional *ad hoc* identification.

5.3. Hypotheses

We test the reduction of the age–period–cohort model to a simpler and more convenient age–cohort model. This is in line with the analysis of Peto *et al.* (1995), although they applied an age–cohort model to mortality rates constructed by using a synthetic measure for exposure. The likelihood ratio or deviance test is defined in Appendix A.2.

Table 1 shows that the deviance of the age–cohort model relative to the age–period–cohort model is 56.8 with $J - 2 = 39$ degrees of freedom, giving a p-value of 0.033. The decision is therefore marginal so the data are not sufficiently informative to tell whether a period effect is needed or not. Thus, from an inferential viewpoint we cannot draw strong conclusions about the period effect. However, from a forecasting viewpoint parsimony is often useful so the period effect will be dropped. The effect of this choice is investigated in a sensitivity analysis in Section 5.6.

The second set of hypotheses concerns the concavity of the age effect and the cohort effect. If these effects were quadratic there would be scope for constructing a very parsimonious model. Indeed, the age effect would be quadratic if $\Delta^2\alpha_3 = \ldots = \Delta^2\alpha_I$, which is a linear hypothesis on the canonical parameter ξ. The relative deviance compared with the unrestricted age–period–cohort model is 228 with $I - 2 = 63$ degrees of freedom. Similarly, the hypothesis of quadratic cohort effects has a relative deviance of 487 with $K - 2 = 103$ degrees of freedom. In both cases the p-value is negligible, which suggests that these effects are more complicated than simple quadratic functions.

Fig. 5 shows the estimates for the age–cohort model. Figs 5(a) and 5(c) show the differences of age and cohort effects. Similarly to the log-odds-ratio interpretation of the double difference these single differences are interpreted as log-odds. The sparsity for the youngest age groups and for the youngest and oldest cohorts results in volatility as before. Fig. 5(a) shows that for older age groups the log-odds are increasing with age.

Figs 5(b) and 5(d) show sums of the differences. In particular, the age effect appears concave in Fig. 5(b). This concavity was also seen in Fig. 4(b), although there it was masked by the unidentifiable linear trend.

5.4. Forecasts from the age–cohort model

Here we describe our forecasting results from the preferred age–cohort model. The first issue is to decide the forecast horizon. The data are an age–period array of dimension $I \times J$, corresponding to ages 25–89 years and periods 1967–2007. For the last period in the sample, i.e. the year 2007, the cohorts are $k = K - I + 1, \ldots, K$, corresponding to the cohorts 1918–1982. The forecasts will be for the next $h = 40$ periods, until 2047, but only for those cohorts that are actually in the sample, giving a forecast index set $\mathcal{J}$ of the form (4.17).

The considerable cohort effects in the data need to be taken into account. The sample includes

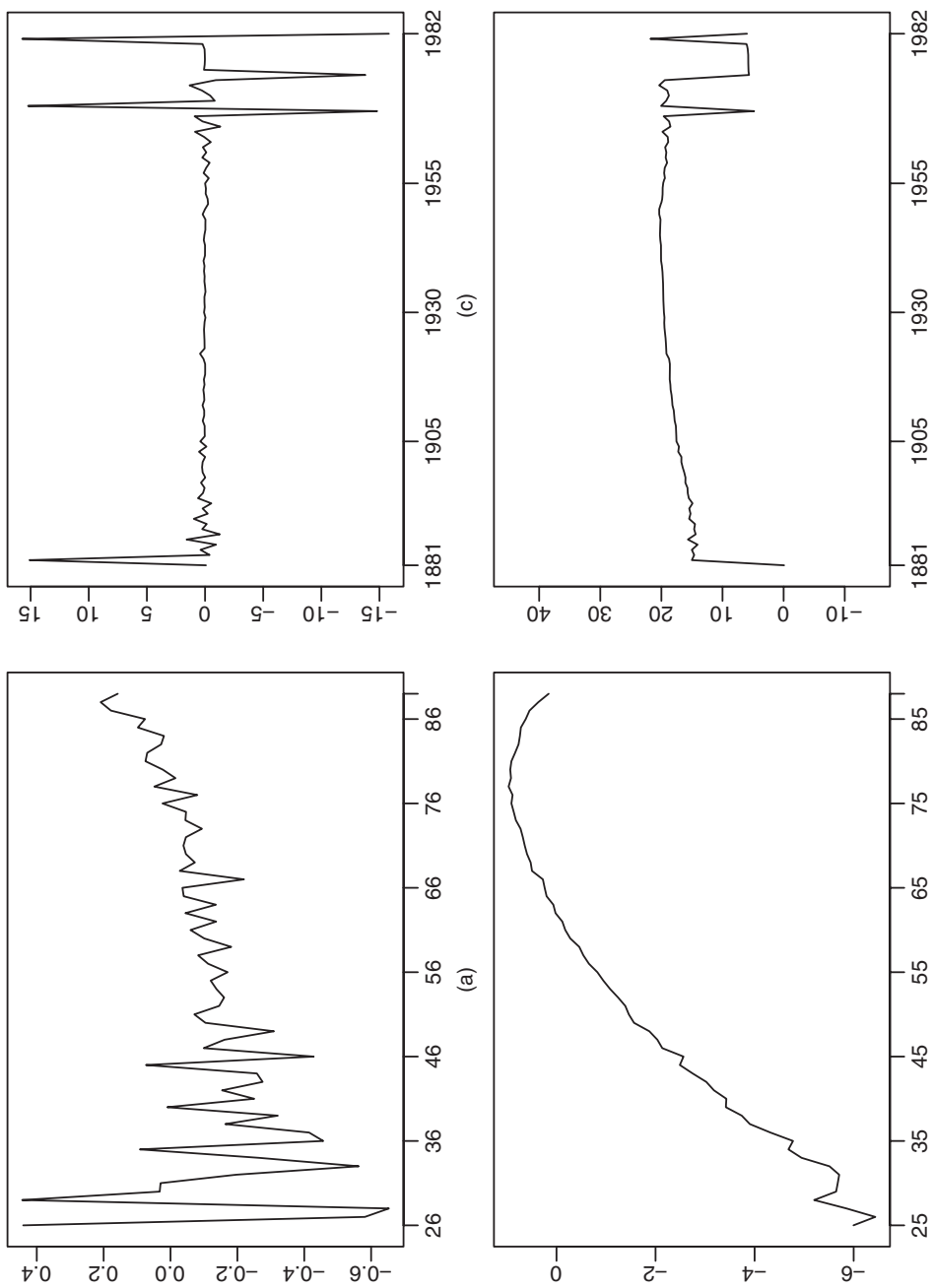

Fig. 5. (a), (b) Estimated age and (c), (d) cohort effects by using two parameterizations, namely (a), (c) the first-order differences of the effects and (b), (d) the sums of these differences

cohorts until 1982. The numbers of observations are sparse for the most recent of these cohorts, 1967–1982 say, for two reasons. First, these cohort groups contain men who are are too young in the sample to have generated many observations as mesothelioma seems to have a long latency. A consequence of this is that the estimates of the parameters for the youngest cohorts will be very uncertain. Secondly, and more importantly, these cohorts have largely been spared from asbestos exposure. Peto *et al.* (1995) explained that the first asbestos regulation was introduced in the UK in 1969. If this regulation has worked as intended the exposure to asbestos will have been modest for these cohorts. It is therefore of interest to break down the forecasts by cohort groups. The cut-off of 1966 is chosen because the 1967 cohort is the first cohort at the end of the mesothelioma epidemic for which no deaths are recorded.

Fig. 6 shows the first forecast results broken down by cohort groups. The dots indicate the observed counts of mesothelioma deaths by period. The central lines are point forecasts with distribution forecasts drawn around them. The top curve represents forecasts of the total number of deaths among those cohorts in which the men were born in 1966 and before. The next curve includes cohorts until 1952 and the bottom curve cohorts until 1937. Forecasts of all cohort groups until 1982 are presented in Fig. 7 and will be discussed later. The forecasts indicate that the number of deaths has already peaked for the oldest cohorts up to 1937. In contrast, the number of deaths for cohorts up to 1952 and 1966 will peak in the future.

Fig. 6 also shows 95% forecast confidence bands. Two bands are shown. The inner bands represent the forecast innovation error only. The outer bands include a contribution from the estimation error, i.e. $s^2_{i,J+h,\text{est}}$ in equation (A.8) in Appendix A. For the two bottom curves representing the oldest cohorts up to 1937 and up to 1952 the estimation error is negligible throughout the forecast and the darker outer band has the appearance of a border to the lighter inner band. For the top curve including cohorts until 1966 the importance of the estimation

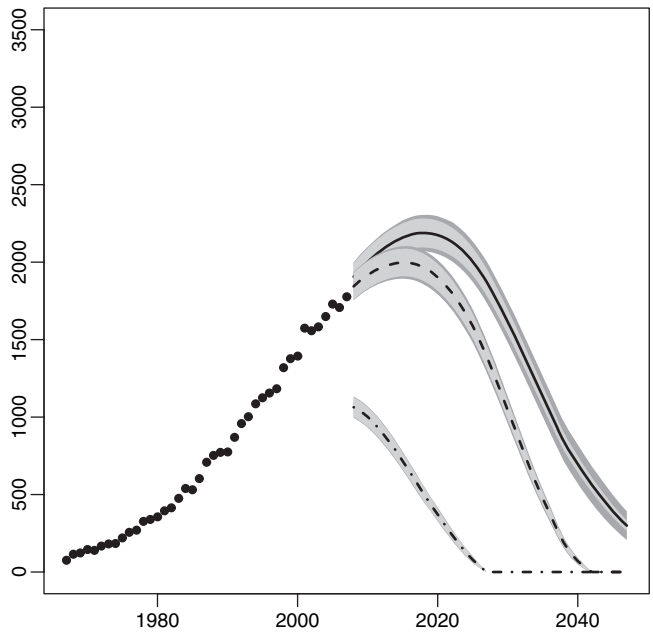

Fig. 6. Forecasts of annual numbers of deaths based on the full sample and decomposed by cohort contribution (age–cohort model): ———, cohorts up to 1966; – – –, cohorts up to 1952; ·–·–·, cohorts up to 1937

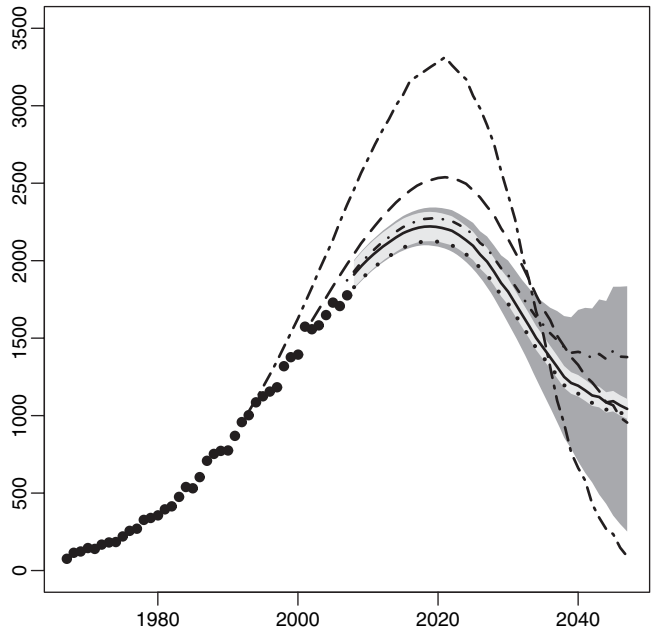

Fig. 7. Recursive forecasts and forecasts of annual number of deaths (age–cohort model): — · —, 1967–1991; — —, 1967–2001; · – · – ·, 1967–2006; ———, 1967–2007; · · · · · ·, 1967–2007, intercept corrected

Table 2. Peaks from recursive analysis

Sample end	*Peak*	*Peak year*
1991	3313	2021
2001	2539	2021
2006	2275	2020
2007	2220	2019
2007†	2125	2019

†Intercept corrected.

error is negligible for the first 10 periods out of sample, whereas it gradually dominates the forecast innovation error. The issue is that at the longest forecast horizons the youngest cohorts will be dominant; indeed the forecast for 2047 will only include cohorts 1958–1966 for which the data are sparse. If all cohorts until 1982 are included this effect will become extreme, as shown for the curve marked '1967–2007' in Fig. 7. For all curves the contribution from the estimation error is modest at the peak in 2019 (Table 2), but it increases rapidly for longer horizons.

5.5. Recursive analysis

Fig. 7 shows the results from a recursive analysis with peak values summarized in Table 2. The purpose of the recursive analysis is to be able to compare the results with those of previous studies. The graph of the observations indicates that after 1991 the increase in mortality is

reduced. This feature can have a large effect on forecast models as discussed in Section 4.3.4 and should be taken into account in retrospective analysis.

The graph marked '1967–1991' uses only the sample until 1991 for the estimation. These are virtually the data that were available to Peto *et al.* (1995). In line with their finding the peak is high, possibly because the 1969 legislation has not yet had an effect due to the long latency of mesothelioma.

The graph marked '1967–2001' uses the sample until 2001 for the estimation. These are virtually the data that were available to Hodgson *et al.* (2005). The forecast is much smaller in line with their finding. Thus, the difference in forecast between Peto *et al.* (1995) and Hodgson *et al.* (2005) appears to be due mainly to the change in mortality and not so much because of the changed method.

Likewise the graph marked '1967–2006' uses the sample corresponding to that of Tan and Warren (2009). It shows a further reduction in the peak in line with the finding of Tan and Warren (2009), who produced a similar recursive analysis.

The graph marked '1967–2007' is based on the full sample and is reported with confidence bands. This has been discussed previously. The graph marked '1967–2007, intercept corrected' is similar but has been subjected to intercept correction of the form (4.28). Thus the difference between the two curves is a scaling factor of $2125/2220 = 0.96$. It is seen how the intercept correction joins up the data and the forecast in a smoother fashion. The intercept correction is, however, relatively modest and stays within the innovation forecast error bands. The choice between these two graphs will depend on whether the recent drop in mortality is seen to be permanent or not.

Overall, the present analysis suggests that the peak will be slightly higher and slightly later than the prediction from previous studies.

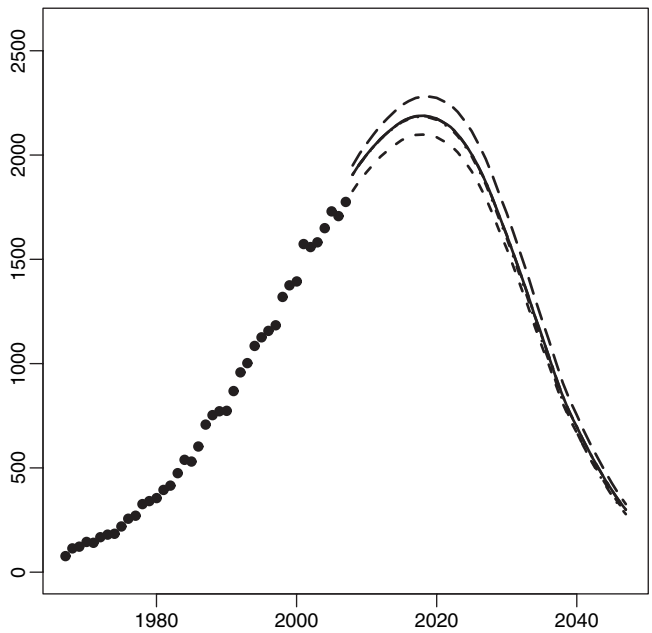

Fig. 8. Sensitivity analysis: forecasts of annual numbers of deaths for cohorts until 1966: — —, age–period–cohort model; ———, age–cohort model; ·–·–·, age–cohort model (35–89 years); - – -, age–cohort model (1878–1966); – – –, age–cohort model, intercept corrected

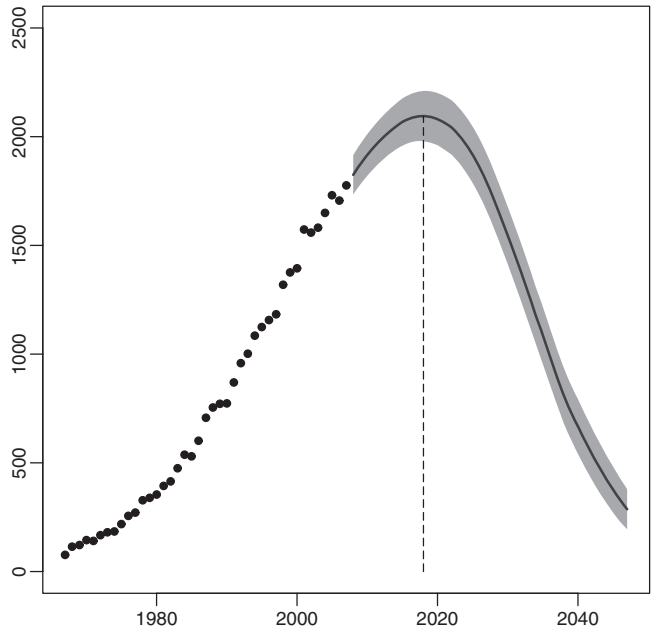

Fig. 9. Forecasts of annual numbers of mesothelioma deaths based on the full sample from the age–cohort model suggested (peak, 2094 deaths in 2018): ▭, pointwise 95% forecast bands

5.6. Sensitivity analysis

Fig. 8 shows how sensitive the results are to variations in the forecasting model. Forecasts are done for cohorts until 1966 as in Fig. 6. Five variations are considered.

The three curves in the middle are nearly identical. One of these is the age–cohort forecast from Fig. 6. The other two are based on age–cohort analyses estimated on reduced samples leaving out either ages 25–34 years or cohorts 1967–1982. This shows that even though the parameters are very poorly estimated for those age and cohort groups this does not contaminate the results.

The bottom curve is based on a full sample age–cohort model. Following the discussion in Section 5.5 the forecast is subjected to an intercept correction. Only cohorts until 1966 are forecasted as for the other graphs in Fig. 8. We take this is as our preferred forecast and discuss the forecast below in Section 5.7.

The top curve is a full sample age–period–cohort forecast. It has the same shape as the age–cohort forecasts but seems less influenced by the drop in mortality in the years 2006–2007. Again, an intercept correction would bring the age–period–cohort forecast in line with the age–cohort forecasts.

5.7. The preferred forecast

On the basis of the above analysis our preferred forecast is the bottom curve from Section 5.5, which is based on an age–cohort model applied to the full sample. We forecast mortality by using only cohorts until 1966. In the forecast an intercept correction is used.

Fig. 9 shows the preferred forecasts along with pointwise 95% confidence interval including both forecast innovation error and estimation error. The analysis points towards a peak in 2018 of about 2094 with 95% confidence interval (1978, 2210). The timing and size of the peak

are broadly in line with the forecasts of Tan *et al.* (2010) and Tan and Warren (2009). But in contrast with them our forecasts have been derived without involving any complicated method to construct a measure of exposure.

This choice of preferred forecast is based on several considerations. First, when testing the age–cohort model against the age–period–cohort model the evidence was inconclusive. In line with this finding the forecasts from the two models are similar; see Fig. 8. On the ground that parsimony often is useful in forecasting, we settle for the age–cohort model. Secondly, cohorts after 1966 have been excluded since the data are very sparse and since these cohorts have been helped by the 1969 legislation. This is rather convenient, since inclusion of these cohorts would lead to a very noisy forecast as shown in Fig. 7. Thirdly, the recursive analysis that is also shown in Fig. 7 indicates that mortality is gradually slowing down. Since forecasting is based on extrapolating the model out of sample this must be taken into account. There are various options for making forecasts robust and we settled for the relatively modest option of an intercept correction.

6. Conclusion

The usual approach to mortality analysis is to model the mortality rates. Mesothelioma mortality is complicated to model in this way because no reliable measure for the exposure exists. Many contributions in the recent literature use sophisticated modelling techniques, with the risk that such detailed modelling increases the modelling error, leading to worse projections.

The purpose of this paper has therefore been to provide a simple benchmark method that should be helpful when checking the robustness of other more sophisticated methods. It has been argued that in so far as interest focuses on predicting overall mortality then the problem can be analysed by using a log-linear model with an age–period–cohort structure, but no offset.

To carry this out some methodological contributions have been made. First, the identification problem of age–period–cohort models has been analysed for age–period arrays. Secondly, it has been shown how to conduct inference by using a multinomial sampling scheme. Thirdly, it has been discussed how to make point forecasts when the period effect must be extrapolated. These contributions are relatively easy to implement by using for instance R which has a routine for generalized linear modelling. The fourth contribution is the discussion of distribution forecasts. This is slightly more complicated to compute.

In the empirical analysis it was found that mesothelioma mortality is expected to peak with about 2094 deaths in 2018, which is slightly worse and slightly later than the predictions of previous studies. Various sensitivity and robustness analyses were carried out. On the basis of these considerations we find that the most appropriate forecast is to apply an age–cohort model and to use this to forecast mortality for cohorts earlier than 1967 by using an intercept correction. This is what was reported in Fig. 9.

Acknowledgements

The data are a minor update of the data that were used by the UK Asbestos Working Party (2009). All numerical computations were done with R; see R Development Core Team (2011). M. D. Martínez-Miranda is supported by the European Commission's Marie Curie Intra-European Fellowship FP7-PEOPLE-2011-IEF project 302600 and the Ministerio de Ciencia e Innovación project MTM2008-03010. B. Nielsen is supported by the Institute for Economic Modelling at the Oxford Martin School.

Appendix A: Details of the mathematical analysis

A.1. Identification

Here, we prove theorem 1. Corollary 2 of Kuang *et al.* (2008a) shows that the parameter ξ in equation (4.2) uniquely identifies the predictor μ_{ij} for an age–period data array. It is therefore left to show formula (4.3).

Rewrite β_j and γ_k by using telescopic sums of the form $\beta_j = \beta_1 + \Sigma_{t=2}^{j} \Delta\beta_t$ and $\Delta\beta_t = \Delta\beta_2 + \Sigma_{s=3}^{t} \Delta^2\beta_s$, with the convention that empty sums are 0, to obtain

$$\beta_j = \beta_1 + (j-1)\,\Delta\beta_2 + \sum_{t=3}^{j} \sum_{s=3}^{t} \Delta^2\beta_s,$$

$$\gamma_k = \gamma_1 + (k-1)\,\Delta\gamma_2 + \sum_{t=3}^{k} \sum_{s=3}^{t} \Delta^2\gamma_s.$$

For the age effect use $\alpha_i = \alpha_I - \Sigma_{t=i+1}^{I} \Delta\alpha_t$ and $\Delta\alpha_t = \Delta\alpha_I - \Sigma_{s=t+1}^{I} \Delta^2\alpha_s$ so that

$$\alpha_i = \alpha_I - (I-i)\,\Delta\alpha_I + \sum_{t=i}^{I-2} \sum_{s=t}^{I-2} \Delta^2\alpha_{s+2}.$$

Substitute these expressions into equation (3.4) noting that $k = I - i + j$ and $\mu_{I1} - \mu_{I-1,1} = \Delta\alpha_I - \Delta\gamma_2$ and $\mu_{I2} - \mu_{I1} = \Delta\beta_2 + \Delta\gamma_2$ to obtain that

$$\mu_{ij} = \mu_{I1} + (i-I)(\mu_{I1} - \mu_{I-1,1}) + (j-1)(\mu_{I2} - \mu_{I1}) + \sum_{t=i}^{I-2} \sum_{s=t}^{I-2} \Delta^2\alpha_{s+2} + \sum_{t=3}^{j} \sum_{s=3}^{t} \Delta^2\beta_s + \sum_{t=3}^{k} \sum_{s=3}^{t} \Delta^2\gamma_s,$$

which is the desired formula (4.3). This concludes the proof of theorem 1.

An alternative parameterization in terms of double sums of double differences arises as follows. Formula (4.3) implies that $\mu = X'_{ij}\xi$, where ξ and X_{ij} are given in equations (4.2) and (4.14). Any one-to-one mapping of ξ will also identify the likelihood. In practice it may be convenient to work directly with the double sums of the double differences, so that $\mu_{ij} = \breve{X}'_{i,j}\breve{\xi}$, where

$$\breve{\xi} = \left(\mu_{I1}, \mu_{I1} - \mu_{I-1,1}, \mu_{I2} - \mu_{I1}, \sum_{t=3}^{I-2} \sum_{s=t}^{I-2} \Delta^2\alpha_{s+2}, \ldots, \sum_{t=I-2}^{I-2} \sum_{s=t}^{I-2} \Delta^2\alpha_{s+2}, \sum_{t=3}^{3} \sum_{s=3}^{t} \Delta^2\beta_s, \ldots, \sum_{t=3}^{J} \sum_{s=3}^{t} \Delta^2\beta_s, \right.$$
$$\left. \sum_{t=3}^{3} \sum_{s=3}^{t} \Delta^2\gamma_s, \ldots, \sum_{t=3}^{K} \sum_{s=3}^{t} \Delta^2\gamma_s \right)', \tag{A.1}$$

and $\breve{X}$ is defined by replacing the function $h(t,s)$ with $\breve{h}(t,s) = \mathbf{1}_{(t=s)}$ in equation (4.14).

A.2. Inference using multinomial sampling

Asymptotic distribution theory for the estimator $\hat{\xi}^{(2)}$ can be established from the multinomial sampling scheme. Note first that

$$\left(\frac{\partial}{\partial \xi^{(2)}} \pi_{ij} \right)' = \pi_{ij} H_{ij}^{(2)} \qquad H_{ij}^{(2)} = X_{ij}^{(2)} - \sum_{s,\, t \in \mathcal{I}} \pi_{st} X_{st}^{(2)}. \tag{A.2}$$

The information for the multinomial likelihood is then

$$i(\xi^{(2)}) = -\frac{\partial^2}{\partial \xi^{(2)\prime}\, \partial \xi^{(2)}} \log\{L(\xi^{(2)}; Y|Y_{..})\} = \hat{\tau}\, i_1(\xi^{(2)}),$$

where the information about one observation is

$$i_1(\xi^{(2)}) = \sum_{i,j} \pi_{ij} H_{ij}^{(2)} X_{ij}^{(2)\prime} = \sum_{i,j} \pi_{ij} H_{ij}^{(2)} H_{ij}^{(2)\prime}. \tag{A.3}$$

Using the δ-method the estimated frequencies are seen to satisfy

$$\tau^{1/2}(\hat{\pi}_{ij} - \pi_{ij}) \xrightarrow{\mathrm{D}} N\{0, \pi_{ij}^2 H_{ij}^{(2)\prime} i_1(\xi^{(2)})^{-1} H_{ij}^{(2)}\}. \tag{A.4}$$

For implementation it is useful to note that the information $i(\xi^{(2)})$ for the multinomial likelihood (4.15) can be obtained from the information for the original likelihood (4.13) as $I_{22\cdot 1} = I_{22} - I_{21} I_{11}^{-1} I_{12}$ where

$$I_{ab} = -\frac{\partial^2}{\partial \xi^{(a)\prime}\, \partial \xi^{(b)}} \log\{L(\xi; Y)\}. \tag{A.5}$$

The goodness of fit is assessed by computing the deviance to the saturated model

$$D(\hat{\xi}; Y) = 2\left[\sum_{i,j} \{Y_{ij} \log(Y_{ij}) - Y_{ij}\} - \sum_{i,j} \{Y_{ij} X'_{ij} \hat{\xi} - \exp(X'_{ij}\hat{\xi})\}\right].$$

It holds for $Y_{\cdot\cdot} \to \infty$ that

$$D(\hat{\xi}; Y) \xrightarrow{\mathrm{D}} \chi^2_{IJ-p}. \tag{A.6}$$

Hypotheses on $\xi^{(2)}$ can likewise be tested by using χ^2-inference. The hypothesis of absence of a period effect is of particular interest and is discussed in what follows.

In Section A.1 it was suggested that the canonical parameter could either be chosen as ξ from equation (4.2) or as $\breve{\xi}$ from equation (A.1), which are equivalent parameterizations. In an empirical study both can be used depending on which aspect of the parameters it is desirable to illustrate. The asymptotic analysis that is discussed here carries over from ξ to $\breve{\xi}$ in a straightforward way because the corresponding design matrices share the first co-ordinate which is used to define τ.

A.3. Distribution forecasting for an age–cohort model

In Section 4.3.3 distribution forecasts were discussed for the age–cohort model. These are of the form

$$\tilde{Y}^{\text{distribution}}_{i,J+h} = \hat{\tau}\tilde{\pi}_{i,J+h} + N(0, s^2_{i,J+h}), \tag{A.7}$$

$$s^2_{i,J+h} = \hat{\tau}\tilde{\pi}_{i,J+h}(1 + s^2_{i,J+h,\text{est}}), \tag{A.8}$$

The term $s^2_{i,J+h,\text{est}}$ arises from the estimation error. It has a form that is similar to that of the estimation uncertainty for $\tilde{\pi}_{ij}$ when $i, j \in \mathcal{I}$ as given in equation (A.4) and it is given by

$$s^2_{i,J+h,\text{est}} = \tilde{\pi}_{i,J+h}\, H^{(2)\prime}_{i,J+h}\, i_1(\hat{\xi}^{(2)})^{-1} H^{(2)}_{i,J+h}. \tag{A.9}$$

Here $\hat{\tau} = Y_{\cdot\cdot}$ is the total number of deaths, $\tilde{\pi}_{i,J+h}$ is the point forecast in equation (4.21), $H^{(2)}_{ij}$ is defined in equation (A.2) and the information $i_1(\hat{\xi}^{(2)})$ is computed as in equation (A.3).

The forecasts can be aggregated over, for instance, age i as follows. The aggregate point forecast is $\tilde{Y}^{\text{point}}_{\cdot,J+h} = \sum_{i=h+1}^{I} \tilde{Y}^{\text{point}}_{i,J+h}$. The innovation errors are independent across cells whereas the estimation errors are dependent across cells. It follows that

$$\tilde{Y}^{\text{point}}_{\cdot,J+h} = \hat{\tau} \sum_{i=h+1}^{I} \tilde{\pi}_{i,J+h}, \tag{A.10}$$

$$\tilde{Y}^{\text{distribution}}_{\cdot,J+h} = \tilde{Y}^{\text{point}}_{\cdot,J+h} + N(0, s^2_{\cdot,J+h}), \tag{A.11}$$

$$s^2_{\cdot,J+h} = \hat{\tau} \sum_{i=h+1}^{I} \tilde{\pi}_{i,J+h} + \hat{\tau} \sum_{s=h+1}^{I} \sum_{t=h+1}^{I} \tilde{\pi}_{s,J+h} H^{(2)\prime}_{s,J+h} i_1^{-1} H^{(2)}_{t,J+h} \tilde{\pi}_{t,J+h}. \tag{A.12}$$

References

Barndorff-Nielsen, O. E. (1978) *Information and Exponential Families.* New York: Wiley.

Berzuini, C. and Clayton, D. (1994) Bayesian analysis of survival on multiple time scales. *Statist Med.*, **13**, 823–838.

Clayton, D. and Schifflers, E. (1987) Models for temporal variation in cancer rates: II, Age-period-cohort models. *Statist Med.*, **6**, 469–481.

Clements, M., Berry, G., Shi, J., Ware, S., Yates, D. and Johnson, A. (2007) Projected mesothelioma incidence in men in New South Wales. *Occupnl Environ. Med.*, **64**, 747–752.

Clements, M. P. and Hendry, D. F. (1999) *Forecasting Non-stationary Time Series*. Cambridge: MIT Press.

Elkum, N. B. (2005) Predicting confidence intervals for the age-period-cohort model. *J. Data Sci.*, **3**, 403–414.

England, P. D. and Verrall, R. J. (2002) Stochastic claims reserving in general insurance. *Br. Act. J.*, **8**, 443–518.

Gardner, M. J. and Osmond, C. (1984) Interpretation of time trends in disease rates in the presence of generation effects. *Statist. Med.*, **3**, 113–130.

Hendry, D. F. and Nielsen, B. (2007) *Econometric Modeling*. Princeton: Princeton University Press.

Hodgson, J. T., McElvenny, D. M., Darnton, A. J., Price, M. J. and Peto, J. (2005) The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. *Br. J. Cancer*, **92**, 587–593.
Holford, T. R. (1985) An alternative approach to statistical age-period-cohort analysis. *J. Chron. Dis.*, **38**, 831–836.
Keiding, N. (1990) Statistical inference in the Lexis diagram. *Phil. Trans, R. Soc. Lond.*, **332**, 487–509.
Kuang, D., Nielsen, B. and Nielsen, J. P. (2008a) Identification of the age-period-cohort model and the extended chain ladder model. *Biometrika*, **95**, 979–986.
Kuang, D., Nielsen, B. and Nielsen, J. P. (2008b) Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika*, **95**, 987–991.
Kuang, D., Nielsen, B. and Nielsen, J. P. (2009) Chain ladder as maximum likelihood revisited. *Ann. Act. Sci.*, **4**, 105–121.
Kuang, D., Nielsen, B. and Nielsen, J. P. (2011) Forecasting in an extended chain-ladder-type model. *J. Risk Insur.*, **78**, 345–359.
Lee, W. C. and Lin, R. S. (1996) Autoregressive age-period-cohort models. *Statist Med.*, **15**, 273–281.
McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
Myojin, T., Azuma, K., Okumura, J. and Uchiyama, I. (2012) Future trends of mesothelioma mortality in Japan based on a risk function. *Industrl Hlth*, **50**, 197–204.
Osmond, C. and Gardner, M. J. (1982) Age, period and cohort models applied to cancer mortality rates. *Statist. Med.*, **1**, 245–259.
Park, E. K., Takahashi, K., Jiang, Y., Movahed, M. and Kameda, T. (2012) Elimination of asbestos use and asbestos-related diseases: an unfinished story. *Cancer Sci.*, **103**, 1751–1755.
Peto, J., Matthews, F. E., Hodgson, T. R. and Jones, J. R. (1995) Continuing increase in mesothelioma mortality in Britain. *Lancet*, **345**, 535–539.
Price, B. and Ware, A. (2009) Time trend of mesothelioma incidence in the United States and projection of future cases: an update based on SEER data for 1973 through 2005. *Crit. Rev. Toxicol.*, **39**, 576–588.
R Development Core Team (2011) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Robinson, B. M. (2012) Malignant pleural mesothelioma: an epidemiological perspective. *Ann. Cardthor. Surg.*, **1**, 491–496.
Segura, O., Burdorf, A. and Looman, C. (2003) Update of predictions of mortality from pleural mesothelioma in the Netherlands. *Occupnl Environ. Med.*, **60**, 50–55.
Tan, E. and Warren, N. (2009) Projection of mesothelioma mortality in Great Britain. *Research Report 728*. Health and Safety Executive, London.
Tan, E., Warren, N., Darnton, A. J. and Hodgson, J. T. (2010) Projection of mesothelioma mortality in Britain using Bayesian methods. *Br. J. Cancer*, **103**, 430–436.
UK Asbestos Working Party (2009) General Insurance Convention, session B12. UK Asbestos Working Party, London. (Available from `http://www.actuaries.org.uk/research-and-resources/pages/general-insurance-convention-2009`.)
Vandenbroucke, J. P. and Pearce, N. (2012) Incidence rates in dynamic populations. *Int. J. Epidem.*, **41**, 1472–1479.